



**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2017

Insights from systems biology in physiological studies: learning from context

Imenez Silva, Pedro Henrique ; Melo, Diogo ; de Mendonça, Pedro Omori Ribeiro

Abstract: Systems biology presents an integrated view of biological systems, focusing on the relations between elements, whether functional or evolutionary, and providing a rich framework for the comprehension of life. At the same time, many low-throughput experimental studies are performed without influence from this integrated view, whilst high-throughput experiments use low-throughput results in their validation and interpretation. We propose an inversion in this logic, and ask which benefits could be obtained from a holistic view coming from high-throughput studies and systems biology in particular in interpreting and designing low-throughput experiments. By exploring some key examples from the renal and adrenal physiology, we try to show that network and modularity theory, along with observed patterns of association between elements in a biological system, can have profound effects on our ability to draw meaningful conclusions from experiments.

DOI: <https://doi.org/10.1159/000478648>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-148767>

Journal Article

Published Version



The following work is licensed under a Creative Commons: Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) License.

Originally published at:

Imenez Silva, Pedro Henrique; Melo, Diogo; de Mendonça, Pedro Omori Ribeiro (2017). Insights from systems biology in physiological studies: learning from context. *Cellular Physiology and Biochemistry*, 42(3):939-951.

DOI: <https://doi.org/10.1159/000478648>

Review

Insights from Systems Biology in Physiological Studies: Learning from Context

Pedro Henrique Imenez Silva^{a,d} Diogo Melo^b Pedro Omori Ribeiro de Mendonça^{c,e}

^aDepartamento de Fisiologia e Biofísica, Instituto de Ciências Biomédicas - Universidade de São Paulo, São Paulo, ^bDepartamento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de São Paulo, São Paulo, ^cDepartamento de Anatomia - Instituto de Ciências Biomédicas - Universidade de São Paulo, São Paulo, Brazil; ^dInstitute of Physiology – Universität Zürich, Zurich, Switzerland; ^eHospital Israelita Albert Einstein, São Paulo, Brazil

Key Words

Physiology • Systems biology • Context • Epistasis • Top-down • Genetic background • Modularity • Interolog

Abstract

Systems biology presents an integrated view of biological systems, focusing on the relations between elements, whether functional or evolutionary, and providing a rich framework for the comprehension of life. At the same time, many low-throughput experimental studies are performed without influence from this integrated view, whilst high-throughput experiments use low-throughput results in their validation and interpretation. We propose an inversion in this logic, and ask which benefits could be obtained from a holistic view coming from high-throughput studies—and systems biology in particular—in interpreting and designing low-throughput experiments. By exploring some key examples from the renal and adrenal physiology, we try to show that network and modularity theory, along with observed patterns of association between elements in a biological system, can have profound effects on our ability to draw meaningful conclusions from experiments.

Introduction

© 2017 The Author(s)
Published by S. Karger AG, Basel

Low-throughput studies in experimental biology are a traditional approach in science and have been responsible for tremendous advances, from new medications to a better understanding of human behaviour. When Otto Loewi, in 1921, showed that acetylcholine was a neurotransmitter, he was using low-throughput techniques to show that one substance could perform a specific action, transmitting a message from one cell to the next [1]. Indeed, research in physiology still uses very similar approaches to identify new molecular pathways, proteins that could be targets for new drugs, mutations in genes that are good candidates for triggering diseases, and so on. Often, low-throughput studies are considered by the academic community as of higher quality when compared to high-throughput

studies. Despite this perception, it has been demonstrated that high-throughput data can be of at least similar quality as data supported by several low-throughput studies [2–4]. On the other hand, the holistic view necessary to compose and interpret high-throughput studies is considered complex, and presents a limiting step to this approach. Modeling and experimentation in high-throughput studies are often labour and technology-intensive, but as more technological restrictions are being overcome, and the quality of the data obtained from high-throughput studies improves, this approach might be, in the future, considered equally important and able to provide high-quality complementary outcomes.

In essence, physiology is a field that deals with biological systems and therefore with the communication between different levels and scales of organization in such environments [5]. Part of the physiology community embraced the high throughput methods as an approach to increase the understanding of those systems. Nevertheless, working with *big data* does not guarantee a higher level of comprehension of a biological question. Systems biology is a field that does not necessarily, but commonly, works with data retrieved from high-throughput approaches, making use of multi-disciplinary groups to interpret them. Physiology has been an important source of knowledge and data for systems biology, but part of the traditional physiology field still ignores what systems biology has to offer. We will argue that systems biology can provide fruitful insights to physiology, and, in this article, we aim to show, using examples from network and modularity theory, along with observed patterns of association between elements in a biological system, that systems biology can have profound effects on our ability to draw meaningful conclusions from experiments, even from traditional fields.

Biological Systems

Regardless of the study model, biological sciences are focused on highly complex systems. As we advance in the study of a biological problem, new components and interactions will be found, or new phenomena will be revealed in different conditions. Scientists deal with this increasing complexity by simplifying or subdividing systems [6]. In order to create a satisfactory simplification of nature, scientists choose a particular feature or perspective of the system to observe and, eventually, to disturb. This perspective defines or implies a scale for the system (like cells, tissues or populations), in which we must be able to perform reliable analyses. At this point, we must decide whether this scale is suitable to answer the posed questions, or if it is necessary to move to another scale. For example, the initial scale in a cell biology context could be the mRNA levels of a group of genes in steady state condition and under a stress condition detected by transcriptome analysis. Then, we are able to move to another scale: the promoter of few altered genes under stress treatment could be examined to find important transcription factor binding sequences, or, in the opposite direction, we could search for organelle malfunctions normally related to the modified transcription profile.

If one were to change scales, this path may go from individual pieces to the whole (bottom-up), or from the whole to the pieces (top-down), and both approaches are used in science. Which one is more common depends on the field, and in some cases even on the university, department, etc. When the direction is bottom-up, the context of a particular study aim is brought into play. In the top-down approach, the context is first visualised and then a path is chosen. In biology, context is essential to understand why a component might behave completely differently in two distinct conditions or environments. This aspect of biological systems was observed in genes, which can act either as tumour promoting or suppressing depending on the context, such as age and type of tumour [7, 8]. The glucuronidase B gene was found to be an oncogene in more aggressive tumours, and a suppressor gene in less aggressive ones [8]. This type of context dependency has been observed even for different cancer subtypes: for example, in breast cancer tumours several suppressor genes might act as promoters depending on the hormonal subtype in which they are expressed [8]. Thus, often a gene allele cannot define the phenotype of presence or absence of a certain tumour

without the wider context. This sort of interaction between different loci leading to different phenotypes, called epistasis, is central to our understanding of how a molecular network is coordinated. Adding to this complexity, phenotypic variability can occur even at the individual cell level in clones (therefore, same set of alleles) under the same drug treatment, which demonstrates that the fate of a cell can be determined by other context based features that have not been identified to date [9].

Genetic Backgrounds

In the next example, we will explore how a gene knockout (KO) in mice can result in completely different phenotypes depending on different genetic backgrounds [10, 11]. The importance of genetic backgrounds and epistasis to understanding the phenotypes of genetically modified mice is illustrated by one example in adrenal gland physiology. Pro-opiomelanocortin (Pomc) is a pro-hormone produced in man mainly by the pituitary gland which, after post-translational processing, results in the production of smaller peptides. These smaller peptides, such as the adrenocorticotrophic hormone and N-terminal peptides, are thought to be essential for steroidogenesis and maintenance of the adrenal gland (for review: [12]). The first mouse model lacking the *Pomc* gene was generated on a mixed background of two mouse lineages, C57BL/6 and 129/SvEv. Interestingly, these mice had no macroscopically discernible adrenal glands [11]. On the other hand, when these mice were backcrossed to a C57BL/6 lineage, thus eliminating confounds due to genetic heterogeneity of a mixed background, the results were surprising: adrenal glands were easily identifiable, however they were hypoplastic as compared to wild-type littermates [10]. How could such a shift in the genetic background lead to dramatically different phenotype? Genetic background can be defined as the genotypes of all other related genes over the genome that may interact with the gene of interest and potentially influence the specific phenotype [13]. As shown in the example above, even a very small amount of other genetic material, mostly from one or two other strains, can significantly influence the final phenotype of a mutant strain. This is probably not because the gene has different functions in different individuals, but because the functional network in these individuals might be different. The presence of a specific allelic combination in each mouse strain could be one of the explanations for these differences. Actually, a particular combination of alleles in the related gene loci of the genetic background can change all the phenotype. This concept can explain the differential susceptibility to a specific disease among humans: the presence of an etiological allele that could result in disease in one subject might not be able to trigger the same phenotype in another (Fig 1A and B). Moreover, it is also possible to explain why the metabolome of a given species changes with a shift in one allele (Fig 1C) [14]. So, the specific allelic combination could generate a particular genetic interaction network that is crucial to understand why the same null mutation shows widely divergent phenotypes depending on the recipient strain.

Unfortunately, it is still difficult to solve such problems, mainly because of two reasons. First, we do not have one database containing all the phenomes found in all mouse strains. Second, until now many of the efforts to characterize these differences in the genetic background are based only in allelic changes and discovery of new single nucleotide polymorphism (SNPs), instead of a holistic view trying to better understand how the genetic networks change between mouse strains. Therefore, if someone were to start a project using one mouse strain instead of another, this could lead to completely different conclusions. One possible way to avoid this problem would be working with congenic strains by a backcross experiment to one of the parental strains. However, this requires precious time and money. It is also highly recommended verifying if the KO model that is studied was already described, by using some available database such as the Mouse Genomic Database (MGD) (<http://www.informatics.jax.org/>) [15], which is an international database resource for laboratory mice, providing integrated genetic, genomic, and biological data to facilitate the study of human health and disease.

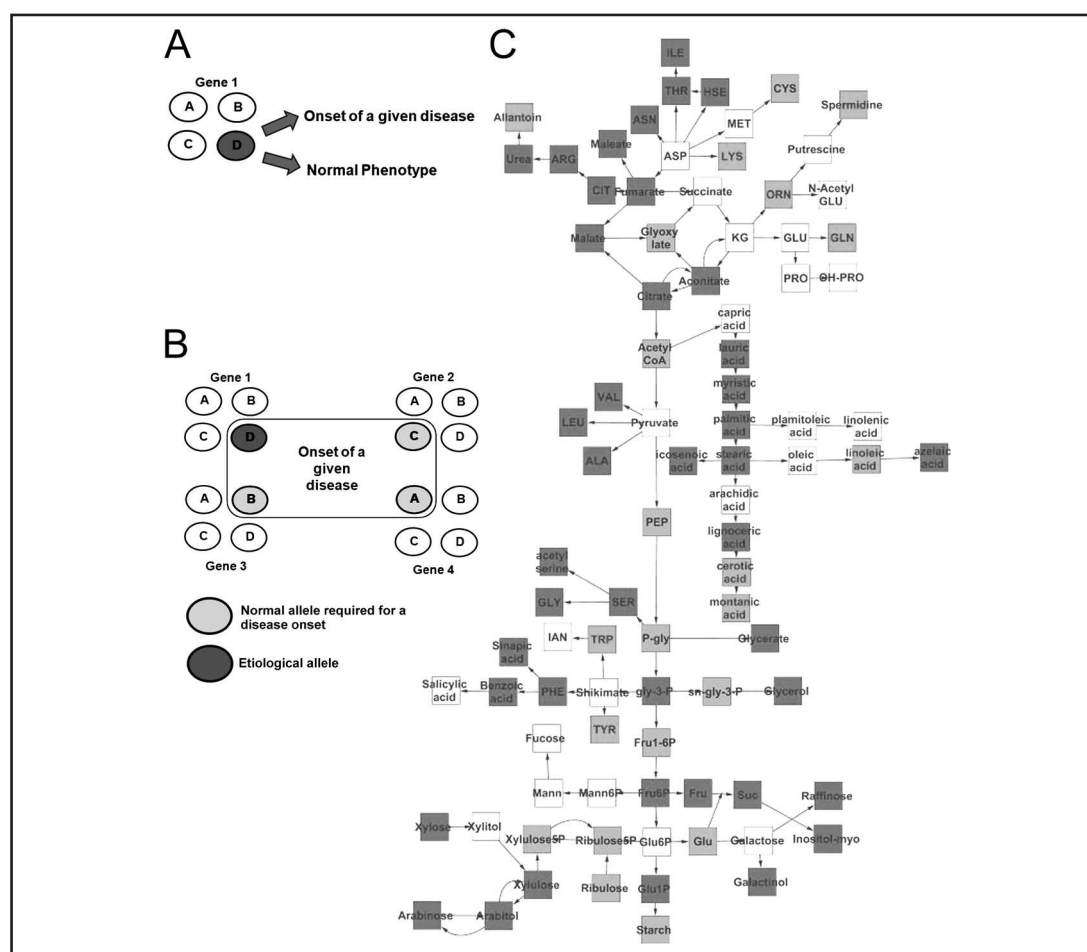


Fig. 1. Two examples of how a shift in alleles can modify the phenotype. (A) The presence of a determined etiological allele does not necessarily lead to disease. (B) The contribution of multiple gene loci (1-4) with multiple alleles (a-d) for the onset of a given disease. Light gray circles represents normal alleles that are necessary for the onset of a disease; Dark gray circles are the etiological alleles for a disease. (C) Genomic variation and metabolome consequences in Arabidopsis. Only one allele (Kas or Tsu) was replaced causing different metabolite accumulation through the central metabolism. Dark gray box – increased metabolite with the Kas allele; Light gray box – increased metabolite with the Tsu allele or metabolite not detected; White box – detected metabolite, but not significantly influenced by the allele replacement. (Fig 1C reprinted according the Creative Commons Attribution License. Original published in [14]).

Thus, changes in genetic networks, possibly evidenced by different backgrounds, might bring about different conclusions from experiments conducted by genetic deletions. However, is it worth all the efforts to generate a complex transgenic animal model to study one gene, even knowing that the gene network in this model probably no longer behaves in an “original” way? Ignoring context may result in conclusions that apply only to specific conditions or even worse, only to the experimental model used. On the other hand, identifying the variables that cause shifts in phenotypes may result in the discovery of whole networks determining complex functions.

From the component to the system

Both bottom-up and top-down approaches can help bring the context into play. Top-down does not necessarily provide a better view of the problem, but at least it guarantees a

context view *a priori*, before the specific experiment is chosen. The bottom-up, as in Plato's cave, allows an outside view of the cave *a posteriori*, revealing the context around it. As top-down provides the context *a priori*, a great advice that systems biology could give would be to choose good targets for low-throughput studies from a top-down approach (for a different perspective see Ellis [16]). As a practical example, it is possible to pick a protein to be analysed, from a proteomic study, and then one or a few proteins are chosen to pass to the next level of studies. Thus, the pipeline goes from a protein system (more complex) to individual proteins (less complex). However, if we discuss how a non-systems biologist could use this approach, if the researcher has an already established laboratory, with established questions being answered, and several students immersed in experiments centred in one or few genes, it is understandable that the advice given above might be ignored. Maybe an alternative advice that could be given would be: why not start from the studied target, go up a level and then return to the initial level to see what can be learned during this journey? We will later discuss some possibilities for this transition.

First, we need to start from lower levels to go to upper levels. As an example, we could set off from the functions of only one gene to understand how a non-systems biology problem may become a systems biology one. In our example, the context in which the gene will be analysed is the kidney. In the proximal tubules, the first segment in the nephrons, 70-80% of the bicarbonate filtered is reabsorbed. Different works show that NHE3, a sodium hydrogen exchanger, is indirectly responsible for 42-65% of this task, by secreting H^+ into the lumen, which forms a HCO_3^- in the cell (Fig 2). These data were retrieved from acute pharmacological, functional and physical-chemical inhibition assays [17–20]. However, only approximately 35% of the HCO_3^- reabsorption was inhibited in proximal tubules of proximal tubule specific NHE3-KO mice [21]. What is the best explanation for this 30% of difference in predicted and observed reabsorption reduction (other than experimental issues)? In all likelihood, there are alternative pathways that can partly compensate the phenotypic perturbation caused by the knockout.

Therefore, it is possible and necessary to look for functional details of the protein local network. Backups can be categorised as genetic buffering or redundancy in the genome at the functional level (e.g. paralogs or isoforms); or through functional complementation (e.g. other proteins with the same function) [22]. In the apical membrane of proximal tubules, there are at least two other NHEs expressed, NHE2 and NHE8, that could provide redundancy to the cells. However, NHE2 does not seem to have a significant participation in H^+ secretion in this segment [23]. NHE8 has also a negligible contribution to this task in steady state conditions, but its expression is up-regulated both in NHE3 KO mice and in wild-type mice subjected to metabolic acidosis [24]. In this system, H^+ -ATPase represents a functional complementation, which accounts for around one third of the indirect HCO_3^- reabsorption by proton secretion in proximal tubules (functional complementation) [25]. Thus, it is expected that some of these backups could substitute NHE3 in case of inhibition or KO, if we take into account *only* the proximal tubule. Since proximal tubular cells have several alternative pathways to secrete protons, even NHE2 might have a role in cells under higher stress conditions or in animals with more than one pathway disrupted (e.g. NHE3 and NHE8 double knockout).

What if some NHE3 function would not show any alteration due to its inhibition of KO? Could we say that this protein is not important in this context? It is assumed that NHE3 is involved in NH_4^+ secretion into urine in the proximal tubule [26], however the same proximal tubule specific NHE3 KO animals cited above have no alteration in urinary ammonium excretion. The authors state in their conclusions that NHE3 is not responsible for this task, despite mentioning that NHE3 might simply not be *alone* in this function. According to the principles stated above, backups for this specific transport could have interfered. Thus, it would be necessary to complement the NHE3 KO approach with pharmacological data, expression data on alternative pathways, and if possible with double KOs (NHE3 + backups) in order to uncover or discard putative NHE3 functions. Double KOs for NHE3/NHE2 and NHE3/NHE8 have been done to analyse impaired proximal tubule HCO_3^- reabsorption, but not

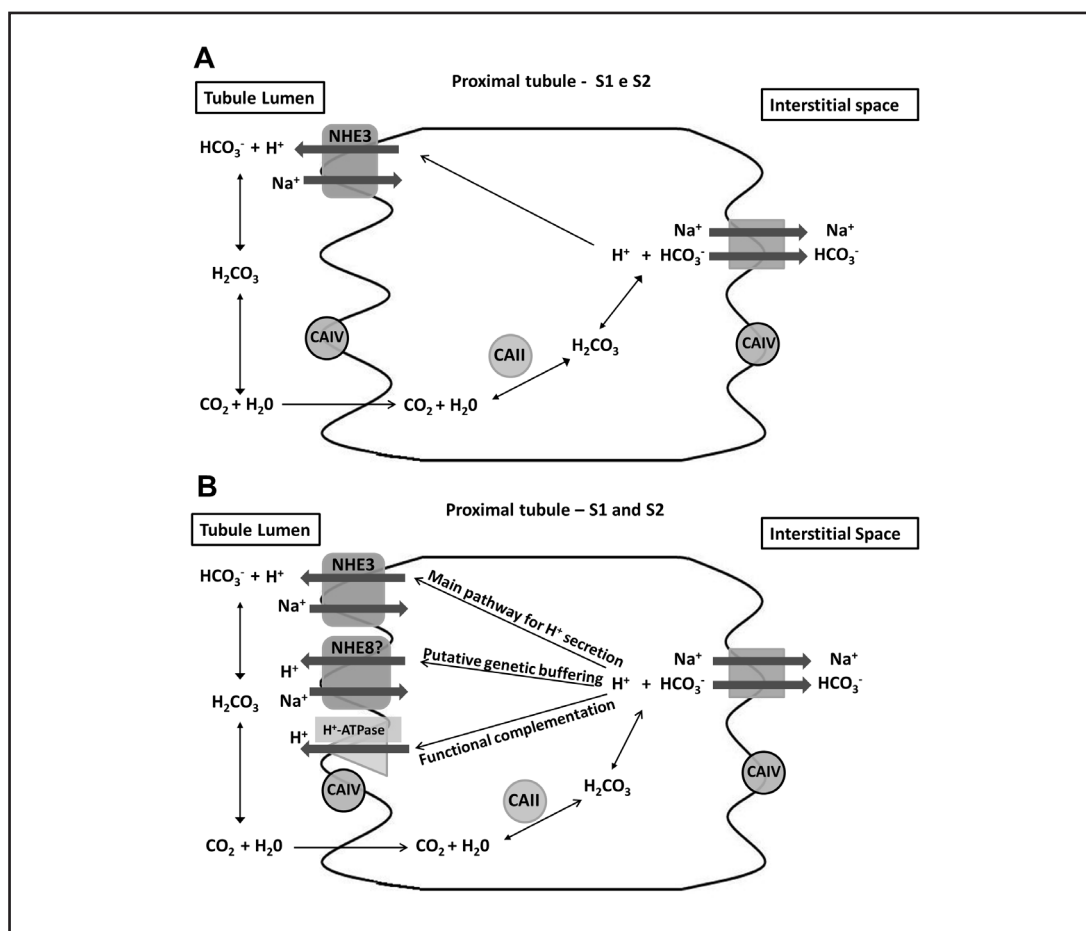


Fig 2. NHE3 backups as an example of adjusting complexity in a local network. The illustrations show how the bicarbonate reabsorption occurs in S1 and S2 proximal tubule segments. Proton secreted by sodium hydrogen exchangers (NHE3 and NHE8) and H^+ -ATPase associates with HCO_3^- in the lumen which creates CO_2 and H_2O . Both molecules enter into the cell and form HCO_3^- and H^+ . The HCO_3^- crosses into the interstitium mainly via a sodium bicarbonate cotransporter (NBC1). (A) H^+ secretion only via NHE3. (B) H^+ secretion including two backups: H^+ -ATPase and NHE8. The figure solely takes in account an increasing in complexity related to the proteins which play the role of proton secretion. Interacting partners and other steps such as the bicarbonate reabsorption in the basolateral membrane could be included. ACII and ACIV mean carbonic anhydrase II and IV, respectively, enzymes that catalyzes the reaction $\text{H}_2\text{O} + \text{CO}_2 \rightleftharpoons \text{HCO}_3^- + \text{H}^+$.

to analyse defects in NH_4^+ secretion [23]. A less cited example is the suggestion of ammonium secretion via luminal potassium channels [27], which reveals that ammonium secretion by the proximal tubule might have its own small network. All these examples around NHE3 take into account only proximal tubule backup mechanisms, but loss of function in one nephron segment is often compensated by gain of function in downstream segments. An alternative approach to avoid data masked by the disturbance provoked by knocking down or knocking in components in a system would be to observe natural variations in the phenotype and correlate them to measurable parameters, such as protein expression. This approach has been used for cell-cell variation in noise genetics [28] and genome-wide association studies [29].

In the previous NHE3 example, functional interactions among components of a system were able to provide suggestions to design experiments and to analyse data. However, there are several components missing in the analysed system. Taking only NHE3 as an example, it physically interacts with other proteins (e.g. CHP, NHERF1 and 2, Megalin, IRBIT, etc.) [30–

33], it receives messages through phosphorylation from other proteins (e.g. CK2, CAMKII, PKA, PKC, etc.) [34–36] and its transcription depends on several transcription factors (e.g. SP1, EGR, etc.) [37–39]. All these interactions are spatially and temporally organized and there are several unknown interactions. It is not possible to incorporate all this information in a simple manner to retrieve functional data. This is especially challenging for highly pleiotropic genes (e.g. Cystic fibrosis transmembrane conductance regulator, CFTR [40, 41]) due to their natural functional complexity and by often exhibiting numerous interaction partners. However, there is an increasing number of data showing new networks at different molecular levels [42]. Thus, one may search for specific components in these databases to hypothesize about its importance in a network, and to design experiments to test for its significance considering the underlying network.

In the *omics* field, gene and protein interaction network databases and tools have been successfully used to predict gene function or genes related to diseases, find targets for new drugs, etc [43–45]. Large and small scale data are compiled by different freely accessible tools that provide a way to find reported or predicted interactions. Some examples are: STRING [46], a database for known and predicted protein-protein interaction (PPI) networks (<http://string-db.org/>); HumanNet, a probabilistic functional gene network for humans [47] (<http://www.functionalnet.org/humannet/>); BioGRID [48], a repository for interaction datasets of genetics, protein and chemical networks (<http://thebiogrid.org/>); HPRD [49], a human protein (interaction) database manually curated, (<http://www.hprd.org/>) and IntAct, a molecular interaction database (<http://www.ebi.ac.uk/intact/>) [50]. Our idea here is not to show an exhaustive list of databases and tools, because there are several possibilities for researchers retrieving data from repositories. However, one should keep in mind that these tools contain data ranging from experimentally validated (e.g. BioGRID) to predicted interactions (e.g. HumanNet). Moreover, interactions occur at different levels from tool to tool, such as physical binary protein-protein interactions and functional interactions among genes. For extensive reviews about biological network tools see [51, 52].

In the NHE3 examples above, we attempted to show *what* are the components of the local systems and then tried to detect the functional relevance on these interactions. In addition to this question, the architecture of a network is informative, showing *how* such components—namely nodes—are organized. Parameters that denote the position of a node in a network are called node centrality parameters. These parameters may provide information that can be used for determining reliable targets in low-throughput experiments. The simplest parameter is the degree centrality, which indicates the number of links that a certain node has in a network. It is conceivable that a node of high degree in a protein-protein interaction network (i.e. a protein with high number of interactions) could have a higher probability of being essential (i.e. is lethal when disrupted) than a protein with a lower degree. However, this is only one parameter, and the correlation is often weak. Thus, other centrality parameters should be taken into consideration such as, betweenness (the number of times a node is the shortest path between all the pair of nodes) or closeness (average shortest path between a node and all the other nodes) [53]. There are several other centrality parameters, for further introduction about this theme see [53, 54]. The default version of the software Cytoscape [55] provides the user friendly plugin NetworkAnalyzer [56], which allows calculating centrality parameters in a given network.

Modularity

Other topological characteristic of interaction networks have been suggested as good predictors of how important a protein is in a given system. These network characteristics usually come in the guise of some sort of modularity theory. Here, modularity refers to the organization of any sort of interactions between units of the system under consideration. For example, interactions might mean physical contact between proteins [57]; correlations in the levels of gene expression under different circumstances or in different tissues or individuals

[58]; participation in the same metabolic pathway or involvement in the same functional group, such as respiration or cell division [59]; and many others. Under many different types of interactions, the pattern of the graph defined by the presence or absence of interactions shows modularity, that is, a distribution of interactions such that some constituents are more related to each other than with other constituents of the system. These highly connected groups are called modules. Modules frequently recover function, even when function was not used to define them. As an example, proteins connected to diseases are not randomly distributed in a map of molecular interactions and form connected subgraphs called disease modules [60]. This suggests that the network topology can carry information relevant for interpreting experimental results. For example, genes within a module tend to have less wide spread mutation effects when compared to genes that connect different modules. These hubs between modules can be thought as integrating edges in networks, and are probably crucial to homeostasis [61, 62]. Furthermore, mutations in hubs between modules tend to be more deleterious than in peripheral genes, which explains why heritable diseases are more frequently caused by mutations in functionally and topologically peripheral genes: the probability of transmitting to the next generation a disease that disrupts a central part of the network is low, as the odds are higher that the subject will not reach the reproductive age [63]. There are also genes that are highly connected within a given module, and these within module hubs can be central in coordinating a set of genes involved in a biological function [64]. The pattern of connections inside modules can also carry information on the function of the individual elements in an interaction network. For example, network motifs (patterns of interaction in directional networks, such as between genes and transcription factors in regulatory networks) tend to repeat themselves in several different networks, and these motifs carry out specific information-processing functions [65]. Identifying motifs and using them to make inference about the function of the different elements in a network can improve our understanding of the whole system, while also providing possible points of experimental intervention. FANMOD (<http://theinf1.informatik.uni-jena.de/motifs/>) and mfinder (<http://www.weizmann.ac.il/mcb/UriAlon/download/network-motif-software>) are mature motif detection softwares that can be used in several types of networks, while MAVisto (<http://mavisto.ipk-gatersleben.de/>) can be used for motif analysis and visualization.

For any given focal protein or gene, we might ask a series of questions related to the specific interactions that provide the context for its function and physiological effects. For example, concerning the module where this gene is included: what other genes are in the same module? What is the functional module size? Which other modules overlap with it? Is the target regulated by or part of a network motif? Even with a target chosen for other reasons, the context, in a very broad sense, can give hints as to alternative pathways or where to look for changes after experimental manipulation. Non-intuitive interactions might be uncovered during the process, as it was observed between the overlapping disease modules of asthma and celiac disease [60]. At the first glance they could be considered totally unrelated diseases, but both are enriched in components of the immunoglobulin A network and even a SNP was found in common between both diseases [66, 67].

This can be extended to more specific network type questions, like: is the target gene/protein highly connected within its module? This can be used as proxy to its importance and centrality within a given physiological function or pathway. Highly connected elements, hubs within a module, are prime candidates for manipulation and may suggest many possibilities for exploring questions. But results are not always restricted to a single functional group, many genes have pleiotropic effects that can span a range of different biological functions (the previously mentioned NHE3 and CFTR are good examples). These effects can be direct or mediated by other genes and proteins, and the network relation between them provides a background to interpret them. Identifying hubs between different modules can lead to important insights.

Interolog

In the previous examples, the spatial context around the studied component was emphasised as essential to understand how the system behaves. However, local networks are often not described in a model species and it may not always be trivial to generate new datasets. Thus, the temporal context is another source of information in this case, because not only paralogs hold reliable information, but orthologs as well. During speciation events, interactions between components, such as protein-protein or DNA-protein are often conserved, which allow to partially compose a network with data from other species. This approach is named in systems biology as interolog [68, 69]. It was originally termed for protein-protein interactions (PPI) and it was expanded to DNA-protein interactions (DPI). Two different approaches are used: only *in silico* new PPI network predictions based on interologs and a systematic production of new networks based on interologs along with PPI network mapping techniques, such as high-throughput yeast two-hybrid [68]. This approach has been used to detect new interactions for vulval development in *C. elegans* from yeast interactions [68], to describe a human pluripotency network from mouse data [70] or to find potential targets for drug and vaccine development against leishmaniasis [71].

Since the amount of PPI and DPI networks databases have recently increased, it is possible to search for orthologs and their interacting partners in these repositories or to access an interolog prediction server, such as BIPS (BIANA Interolog Prediction Server, <http://sbi.imim.es/BIPS.php>) in order to obtain a list of partner candidates for the species of interest [72]. There are online available tools for finding orthologs, such as EggNog and MetaPhOrs [73, 74]. EggNOG is a blast-based approach which produces orthology inferences in a genome wide manner, while MetaPhOrs is a public repository of phylogeny based orthology (and paralogy) which uses other public homology repositories for its predictions. Nevertheless, an interaction in a given species might not be present in another one. There are good attempts to predict if an interolog interaction is reliable, and some protein features, such as identity between orthologs and protein structure can be used to reduce the chance of inferring a non-existent interaction [69, 75], however, it is not possible to completely avoid false positives. Thus, interolog must be viewed as a potential way to create biological hypotheses, but not as a proof of interaction.

Conclusions

Systems biology applied to experimental biology can provide fruitful insights to low-throughput experimentation in biological and biomedical studies. Simple concepts used to retrieve information from the biological context in which the posed question lies may help not to be misguided by simple, but inaccurate, conclusions. The local network around a gene or protein may assist in finding the most important pathways in which a biological function is established, which relies especially on identifying redundancy and interacting partners. From the local network is also possible to understand how the interactions are organized, which in turn allows one to identify genes and proteins of higher susceptibility to network disruption, proteins that could be targeted by drugs, or to understand the relevance of a module for some biological functions. Beyond the spatial context, the temporal context contains reliable information that can provide further insights and often can be the first step in an unexplored space. Generating new databases and looking for better genomic annotation is essential in improving our biological understanding. Nevertheless, we would like to emphasize that there are already abundant and rich resources of information full of hints to be studied *in silico* and on the bench. To obtain and interpret high quality data can be a large challenge and certainly collaborations are necessary to deal with sufficient statistics, modelling and experimental methods.

In summary, going from the target molecule to its evolutionary history, crucial context is revealed. With this contextual information, a top-down approach can be followed, which

may be tortuous, but we stand to gain from this. As discussed above, there are several simple tools to aid in this journey.

Acknowledgments

We thank Guilherme Garcia for his contributions, helpful discussions and comments on the manuscript; also Jason Wolf, Fabio Machado and Carsten A. Wagner for reading draft versions and providing their comments and suggestions.

We apologise for not mentioning several important databases, tools and resources that could be useful for the readers.

PORM is a recipient of Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) grant number 2012/23535-4, DM is a recipient of FAPESP grant number 2014/26262-4. At the time this work was carried out, PHIS was a recipient of Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes) grant number 18389-12-0 and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) grant number 142491/2008-0. PHIS is a recipient of the European Union's Seventh Framework Programme for research, technological development and demonstration under the grant agreement no 608847, Swiss National Foundation and CNPq grant number 205625/2014-2.

Disclosure Statement

The authors have declared no conflict of interest.

References

- Loewi O: Über humorale übertragbarkeit der Herznervenwirkung. *Pflüg Arch Für Gesamte Physiol Menschen Tiere* 1921;189:239–242.
- Arabidopsis Interactome Mapping Consortium: Evidence for network evolution in an Arabidopsis interactome map. *Sci* 80- 2011;333:601–607.
- Braun P: Interactome mapping for analysis of complex phenotypes: Insights from benchmarking binary interaction assays. *Proteomics* 2012;Available from: PM:22589225
- Yu H, Braun P, Yildirim M, Lemmens I, Venkatesan K, Sahalie J, Hirozane-Kishikawa T, Gebreab F, Li N, Simonis N, Hao T, Rual J, Dricot A, Vazquez A, Murray R, Simon C, Tardivo L, Tam S, Svrikapa N, Fan C, de Smet A, Motyl A, Hudson M, Park J, Xin X, Cusick M, Moore T, Boone C, Snyder M, Roth F, Barabasi A, Tavernier J, Hill D, Vidal M: High-quality binary protein interaction map of the yeast interactome network. *Sci* 80- 2008;322:104–110.
- Strange K: The end of “naïve reductionism”: rise of systems biology or renaissance of physiology? *Am J Physiol - Cell Physiol* 2005;288:C968–C974.
- Hayden EC: Human genome at ten: Life is complicated. *Nat News* 2010;464:664–667.
- Bains W: Paradoxes of Non-Trivial Gene Networks: How Cancer-Causing Mutations Can Appear to Be Cancer-Protective. *Rejuvenation Res* 2004;7:199–210.
- de Souza JES, Fonseca AF, Valieris R, Carraro DM, Wang JYJ, Kolodner RD, de Souza SJ: S-Score: A Scoring System for the Identification and Prioritization of Predicted Cancer Genes. *PLoS ONE* 2014;9:e94147.
- Zimmer A, Amar-Farkash S, Danon T, Alon U: Dynamic proteomics reveals bimodal protein dynamics of cancer cells in response to HSP90 inhibitor. *BMC Syst Biol* 2017;11. DOI: 10.1186/s12918-017-0410-8
- Smart JL, Low MJ: Lack of Proopiomelanocortin Peptides Results in Obesity and Defective Adrenal Function but Normal Melanocyte Pigmentation in the Murine C57BL/6 Genetic Background. *Ann N Y Acad Sci* 2003;994:202–210.
- Yaswen L, Diehl N, Brennan MB, Hochgeschwender U: Obesity in the mouse model of pro-opiomelanocortin deficiency responds to peripheral melanocortin. *Nat Med* 1999;5:1066–1070.

- 12 Bicknell AB: The Tissue-Specific Processing of Pro-Opiomelanocortin. *J Neuroendocrinol* 2008;20:692–699.
- 13 Yoshiki A, Moriwaki K: Mouse Phenome Research: Implications of Genetic Background. *ILAR J* 2006;47:94–102.
- 14 Joseph B, Corwin JA, Li B, Atwell S, Kliebenstein DJ: Cytoplasmic genetic variation and extensive cytonuclear interactions influence natural variation in the metabolome. *eLife* 2013;2. DOI: 10.7554/eLife.00776
- 15 Blake JA, Richardson JE, Davisson MT, Eppig JT: The Mouse Genome Database (MGD). A comprehensive public resource of genetic, phenotypic and genomic data. *Nucleic Acids Res* 1997;25:85–91.
- 16 Ellis GFR: Top-down causation and emergence: some comments on mechanisms. *Interface Focus* 2011;rsfs20110062.
- 17 Baum M: Developmental Changes in Rabbit Juxtamedullary Proximal Convoluted Tubule Acidification. *Pediatr Res* 1992;31:411–414.
- 18 Carraro-Lacroix LR, Lessa LMA, Bezerra CNA, Pessoa TD, Souza-Menezes J, Morales MM, Girardi ACC, Malnic G: Role of CFTR and CIC-5 in Modulating Vacuolar H⁺-ATPase Activity in Kidney Proximal Tubule. *Cell Physiol Biochem* 2010;26:563–576.
- 19 Queiroz-Leite GD, Crajoinas RO, Neri EA, Bezerra CNA, Girardi ACC, Rebouças NA, Malnic G: Fructose Acutely Stimulates NHE3 Activity in Kidney Proximal Tubule. *Kidney Blood Press Res* 2012;36:320–334.
- 20 Wang T, Hropot M, Aronson PS, Giebisch G: Role of NHE isoforms in mediating bicarbonate reabsorption along the nephron. *Am J Physiol - Ren Physiol* 2001;281:F1117–F1122.
- 21 Li HC, Du Z, Barone S, Rubera I, McDonough AA, Tauc M, Zahedi K, Wang T, Soleimani M: Proximal tubule specific knockout of the Na⁺/H⁺ exchanger NHE3: effects on bicarbonate absorption and ammonium excretion. *J Mol Med* 2013;91:951–963.
- 22 Stelling J, Sauer U, Szallasi Z, Doyle III FJ, Doyle J: Robustness of Cellular Functions. *Cell* 2004;118:675–685.
- 23 Choi JY, Shah M, Lee MG, Schultheis PJ, Shull GE, Muallem S, Baum M: Novel amiloride-sensitive sodium-dependent proton secretion in the mouse proximal convoluted tubule. *J Clin Invest* 2000;105:1141–1146.
- 24 Baum M, Twombly K, Gattineni J, Joseph C, Wang L, Zhang Q, Dwarakanath V, Moe OW: Proximal tubule Na⁺/H⁺ exchanger activity in adult NHE3^{-/-}, NHE3^{-/-}, and NHE3^{-/-}/NHE8^{-/-} mice. *Am J Physiol - Ren Physiol* 2012;303:F1495–F1502.
- 25 Fiori M, Radrizzani M, Díaz-Sylvester P, Müller A, Corti T, Monserrat A, Amorena C: Relative contribution of V-H⁺ATPase and Na⁺/H⁺ exchanger to bicarbonate reabsorption in proximal convoluted tubules of old rats. *Aging Cell* 2006;5:367–372.
- 26 Weiner ID, Verlander JW: Role of NH₃ and NH₄⁺ transporters in renal acid-base transport. *Am J Physiol - Ren Physiol* 2011;300:F11–F23.
- 27 Simon EE, Merli C, Herndon J, E. J. Cragoe J, Hamm LL: Effects of barium and 5-(N-ethyl-N-isopropyl)-amiloride on proximal tubule ammonia transport. *Am J Physiol - Ren Physiol* 1992;262:F36–F39.
- 28 Farkash-Amar S, Zimmer A, Eden E, Cohen A, Geva-Zatorsky N, Cohen L, Milo R, Sigal A, Danon T, Alon U: Noise Genetics: Inferring Protein Function by Correlating Phenotype with Protein Levels and Localization in Individual Human Cells. *PLOS Genet* 2014;10:e1004176.
- 29 Visscher PM, Brown MA, McCarthy MI, Yang J: Five Years of GWAS Discovery. *Am J Hum Genet* 2012;90:7–24.
- 30 Babich V, Vadnagara K, Sole FD: The biophysical and molecular basis of intracellular pH sensing by Na⁺/H⁺ exchanger-3. *FASEB J* 2013; DOI: 10.1096/fj.12-225466
- 31 Biemesderfer D, Nagy T, DeGray B, Aronson P: Specific association of megalin and the Na⁺/H⁺ exchanger isoform NHE3 in the proximal tubule. *J Biol Chem* 1999;274:17518–17524.
- 32 He P, Zhang H, Yun C: IRBIT, inositol 1,4,5-triphosphate (IP₃) receptor-binding protein released with IP₃, binds Na⁺/H⁺ exchanger NHE3 and activates NHE3 activity in response to calcium. *J Biol Chem* 2008;283:33544–33553.
- 33 Yun C-HC, Lamprecht G, Forster DV, Sidor A: NHE3 Kinase A Regulatory Protein E3KARP Binds the Epithelial Brush Border Na⁺/H⁺ Exchanger NHE3 and the Cytoskeletal Protein Ezrin. *J Biol Chem* 1998;273:25856–25863.
- 34 Azarani A, Goltzman D, Orlowski J: Structurally Diverse N-terminal Peptides of Parathyroid Hormone (PTH) and PTH-related Peptide (PTHrP) Inhibit the Na⁺/H⁺ Exchanger NHE3 Isoform by Binding to the PTH/PTHrP Receptor Type I and Activating Distinct Signaling Pathways. *J Biol Chem* 1996;271:14931–14936.

- 35 Sarker R, Gronborg M, Cha B, Mohan S, Chen Y, Pandey A, Litchfield D, Donowitz M, Li X: Casein kinase 2 binds to the C terminus of Na⁺/H⁺ exchanger 3 (NHE3) and stimulates NHE3 basal activity by phosphorylating a separate site in NHE3. *Mol Biol Cell* 2008;19:3859–3870.
- 36 Zizak M, Chen T, Bartonicek D, Sarker R, Zachos N, Cha B, Kovbasnjuk O, Korac J, Mohan S, Cole R, Chen Y, Tse C, Donowitz M: Calmodulin kinase II constitutively binds, phosphorylates, and inhibits brush border Na⁺/H⁺ exchanger 3 (NHE3) by a NHERF2 protein-dependent process. *J Biol Chem* 2012;287:13442–13456.
- 37 Malakooti J, Sandoval R, Amin M, Clark J, Dudeja P, Ramaswamy K: Transcriptional stimulation of the human NHE3 promoter activity by PMA: PKC independence and involvement of the transcription factor EGR-1. *Biochem J* 2006;396:327–336.
- 38 Neri E, Bezerra C, Reboucas N: Essential regulatory elements for NHE3 gene transcription in renal proximal tubule cells. *Braz J Med Biol Res* 2011;44:514–523.
- 39 Silva PHI, Girardi ACC, Neri EA, Rebouças NA: Distinct mechanisms underlie adaptation of proximal tubule Na⁺/H⁺ exchanger isoform 3 in response to chronic metabolic and respiratory acidosis. *Pflug Arch - Eur J Physiol* 2012;463:703–714.
- 40 Valdivieso ÁG, Clauzure M, Massip-Copiz M, Santa-Coloma TA: The Chloride Anion Acts as a Second Messenger in Mammalian Cells - Modifying the Expression of Specific Genes. *Cell Physiol Biochem* 2016;38:49–64.
- 41 Vankeerberghen A, Cuppens H, Cassiman J-J: The cystic fibrosis transmembrane conductance regulator: an intriguing protein with pleiotropic functions. *J Cyst Fibros* 2002;1:13–29.
- 42 Bhat A, Dakna M, Mischak H: Integrating Proteomics Profiling Data Sets: A Network Perspective; in Vlahou A, Makridakis M (eds): *Clinical Proteomics*. Springer New York, 2015, pp 237–253.
- 43 Miller ML, Molinelli EJ, Nair JS, Sheikh T, Samy R, Jing X, He Q, Korkut A, Crago AM, Singer S, Schwartz GK, Sander C: Drug Synergy Screen and Network Modeling in Dedifferentiated Liposarcoma Identifies CDK4 and IGF1R as Synergistic Drug Targets. *Sci Signal* 2013;6:ra85.
- 44 Mostafavi S, Ray D, Warde-Farley D, Grouios C, Morris Q: GeneMANIA: a real-time multiple association network integration algorithm for predicting gene function. *Genome Biol* 2008;9:S4.
- 45 Özgür A, Vu T, Erkan G, Radev DR: Identifying gene-disease associations using centrality on a literature mined gene-interaction network. *Bioinformatics* 2008;24:i277–i285.
- 46 Mering C von, Huynen M, Jaeggi D, Schmidt S, Bork P, Snel B: STRING: a database of predicted functional associations between proteins. *Nucleic Acids Res* 2003;31:258–261.
- 47 Lee I, Blom UM, Wang PI, Shim JE, Marcotte EM: Prioritizing candidate disease genes by network-based boosting of genome-wide association data. *Genome Res* 2011;21:1109–1121.
- 48 Stark C, Breitkreutz B-J, Reguly T, Boucher L, Breitkreutz A, Tyers M: BioGRID: a general repository for interaction datasets. *Nucleic Acids Res* 2006;34:D535–D539.
- 49 Peri S, Navarro JD, Amanchy R, Kristiansen TZ, Jonnalagadda CK, Surendranath V, Niranjana V, Muthusamy B, Gandhi TKB, Gronborg M, Ibarrola N, Deshpande N, Shanker K, Shivashankar HN, Rashmi BP, Ramya MA, Zhao Z, Chandrika KN, Padma N, Harsha HC, Yatish AJ, Kavitha MP, Menezes M, Choudhury DR, Suresh S, Ghosh N, Saravana R, Chandran S, Krishna S, Joy M, Anand SK, Madavan V, Joseph A, Wong GW, Schiemann WP, Constantinescu SN, Huang L, Khosravi-Far R, Steen H, Tewari M, Ghaffari S, Globe GC, Dang CV, Garcia JGN, Pevsner J, Jensen ON, Roepstorff P, Deshpande KS, Chinnaiyan AM, Hamosh A, Chakravarti A, Pandey A: Development of Human Protein Reference Database as an Initial Platform for Approaching Systems Biology in Humans. *Genome Res* 2003;13:2363–2371.
- 50 Hermjakob H, Montecchi-Palazzi L, Lewington C, Mudali S, Kerrien S, Orchard S, Vingron M, Roehert B, Roepstorff P, Valencia A, Margalit H, Armstrong J, Bairoch A, Cesareni G, Sherman D, Apweiler R: IntAct: an open source molecular interaction database. *Nucleic Acids Res* 2004;32:D452–D455.
- 51 Ghosh S, Matsuoka Y, Asai Y, Hsin K-Y, Kitano H: Software for systems biology: from tools to integrated platforms. *Nat Rev Genet* 2011;12:821–832.
- 52 Ng A, Bursteinas B, Gao Q, Mollison E, Zvelebil M: Resources for integrative systems biology: from data through databases to networks and dynamic system models. *Brief Bioinform* 2006;7:318–330.
- 53 Ma'ayan A: Introduction to Network Analysis in Systems Biology. *Sci Signal* 2011;4:tr5.
- 54 Borgatti SP, Everett MG: A Graph-theoretic perspective on centrality. *Soc Netw* 2006;28:466–484.
- 55 Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T: Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003;13:2498–504.

- 56 Assenov Y, Ramírez F, Schelhorn S-E, Lengauer T, Albrecht M: Computing topological parameters of biological networks. *Bioinformatics* 2008;24:282–284.
- 57 Han J-DJ, Bertin N, Hao T, Goldberg DS, Berriz GF, Zhang LV, Dupuy D, Walhout AJM, Cusick ME, Roth FP, Vidal M: Evidence for dynamically organized modularity in the yeast protein–protein interaction network. *Nature* 2004;430:88–93.
- 58 Ayroles JF, Carbone MA, Stone EA, Jordan KW, Lyman RF, Magwire MM, Rollmann SM, Duncan LH, Lawrence F, Anholt RRH, Mackay TFC: Systems genetics of complex traits in *Drosophila melanogaster*. *Nat Genet* 2009;41:299–307.
- 59 Costanzo M, Baryshnikova A, Bellay J, Kim Y, Spear ED, Sevier CS, Ding H, Koh JLY, Toufighi K, Mostafavi S, Prinz J, St Onge RP, VanderSluis B, Makhnevych T, Vizeacoumar FJ, Alizadeh S, Bahr S, Brost RL, Chen Y, Cokol M, Deshpande R, Li Z, Lin Z-Y, Liang W, Marback M, Paw J, San Luis B-J, Shuteriqi E, Tong AHY, van Dyk N, Wallace IM, Whitney JA, Weirauch MT, Zhong G, Zhu H, Houry WA, Brudno M, Ragibizadeh S, Papp B, PääfÅl C, Roth FP, Giaever G, Nislow C, Troyanskaya OG, Bussey H, Bader GD, Gingras A-C, Morris QD, Kim PM, Kaiser CA, Myers CL, Andrews BJ, Boone C: The genetic landscape of a cell. *Sci* 80- 2010;327:425–31.
- 60 Menche J, Sharma A, Kitsak M, Ghiassian SD, Vidal M, Loscalzo J, Barabási A-L: Uncovering disease-disease relationships through the incomplete interactome. *Science* 2015;347:1257601.
- 61 Fraser HB: Coevolution, modularity and human disease. *Curr Opin Genet Dev* 2006;16:637–644.
- 62 Wagner GP, Pavlicev M, Cheverud JM: The road to modularity. *Nat Rev Genet* 2007;8:921–31.
- 63 Goh K-I, Cusick ME, Valle D, Childs B, Vidal M, Barabási A-L: The human disease network. *Proc Natl Acad Sci* 2007;104:8685–8690.
- 64 Seo CH, Kim J-R, Kim M-S, Cho K-H: Hub genes with positive feedbacks function as master switches in developmental gene regulatory networks. *Bioinformatics* 2009;25:1898–1904.
- 65 Alon U: Network motifs: theory and experimental approaches. *Nat Rev Genet* 2007;8:450–461.
- 66 Ferreira MAR, Matheson MC, Duffy DL, Marks GB, Hui J, Le Souëf P, Danoy P, Baltic S, Nyholt DR, Jenkins M, Hayden C, Willemsen G, Ang W, Kuokkanen M, Beilby J, Cheah F, de Geus EJC, Ramasamy A, Vedantam S, Salomaa V, Madden PA, Heath AC, Hopper JL, Visscher PM, Musk B, Leeder SR, Jarvelin M-R, Pennell C, Boomsma DI, Hirschhorn J, Walters H, Martin NG, James A, Jones G, Abramson MJ, Robertson CF, Dharmage SC, Brown MA, Montgomery GW, Thompson PJ: Identification of IL6R and chromosome 11q13.5 as risk loci for asthma. *Lancet* 2011;378:1006–1014.
- 67 Hunt KA, Zhernakova A, Turner G, Heap GAR, Franke L, Bruinenberg M, Romanos J, Dinesen LC, Ryan AW, Panesar D, Gwilliam R, Takeuchi F, McLaren WM, Holmes GKT, Howdle PD, Walters JRF, Sanders DS, Playford RJ, Trynka G, Mulder CJJ, Mearin ML, Verbeek WHM, Trimble V, Stevens FM, O'Morain C, Kennedy NP, Kelleher D, Pennington DJ, Strachan DP, McArdle WL, Mein CA, Wapenaar MC, Deloukas P, McGinnis R, McManus R, Wijmenga C, van Heel DA: Newly identified genetic risk variants for celiac disease related to the immune response. *Nat Genet* 2008;40:395–402.
- 68 Walhout A, Sordella R, Lu X, Hartley J, Temple G, Brasch M, Thierry-Mieg N, Vidal M: Protein interaction mapping in *C. elegans* using proteins involved in vulval development. *Sci* 80- 2000;287:116–122.
- 69 Yu H, Luscombe NM, Lu HX, Zhu X, Xia Y, Han J-DJ, Bertin N, Chung S, Vidal M, Gerstein M: Annotation Transfer Between Genomes: Protein–Protein Interologs and Protein–DNA Regulogs. *Genome Res* 2004;14:1107–1118.
- 70 Som A, Luštrek M, Singh NK, Fuellen G: Derivation of an interaction/regulation network describing pluripotency in human. *Gene* 2012;502:99–107.
- 71 Rezende AM, Folador EL, Resende D de M, Ruiz JC: Computational Prediction of Protein-Protein Interactions in *Leishmania* Predicted Proteomes. *PLoS ONE* 2012;7:e51304.
- 72 Garcia-Garcia J, Schleker S, Klein-Seetharaman J, Oliva B: BIPS: BIANA Interolog Prediction Server: A tool for protein-protein interaction inference. *Nucleic Acids Res* 2012;40:W147–W151.
- 73 Huerta-Cepas J, Szklarczyk D, Forslund K, Cook H, Heller D, Walter MC, Rattei T, Mende DR, Sunagawa S, Kuhn M, Jensen LJ, von Mering C, Bork P: eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res* 2016;44:D286–D293.
- 74 Pryszcz LP, Huerta-Cepas J, Gabaldón T: MetaPhOrs: orthology and paralogy predictions from multiple phylogenetic evidence using a consistency-based confidence score. *Nucleic Acids Res* 2011;39:e32–e32.
- 75 Chen Y-C, Lo Y-S, Hsu W-C, Yang J-M: 3D-partner: a web server to infer interacting partners and binding models. *Nucleic Acids Res* 2007;35:W561–W567.